

Going Large (Language Models) at ISMB2023

Scott Edmunds 

Published August 3, 2023

Citation

Edmunds, S. (n.d.). In *GigaBlog*. GigaBlog. <https://doi.org/10.59350/59yfe-63e94>

Keywords

Biology, AI, Bioinformatics, BOSC, Conferences
Feature Image

Copyright

Copyright © Scott Edmunds 2023. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Once again the *GigaScience Press* team has gathered at the yearly [ISMB \(Intelligent Systems for Molecular Biology\)](#) meeting to find out about the state of the art of computational biology, as well as celebrate our birthday.

The conference centre in Lyon, for a conference covering subjects including Large Language Models

Hosted this year in the beautiful city of Lyon, and this year collocated with ISMB's European sibling ECCB, it's now been 11 years since *GigaScience* journals launch at [ISMB 2012 in Long Beach](#), and we have [attended every in-person meeting](#) since Vienna in 2011. After skipping two years because of COVID, [last year in Madison](#) was relatively successful at dipping ISMB's toes into offering a hybrid format while still having a reasonable in-person turnout.

The return of the conference

But this year proved that conferences are properly back, with the highest turnout for the meeting to date: 2100 in-person attendees and 400 watching online. It was great to see lots of familiar faces that hadn't been back for many years, and also to see the society regain some of their economic losses from COVID, although if anything the meeting might have been a bit of a victim of its own success as at times the organisational infrastructure was a little stretched from too many people going up the escalators, challenges fitting into some of the sessions (especially at the keynotes), and long queues for coffee and lunch. With many PhD students having missed out on the important experience of attending conferences because of COVID many PI's were generous in bringing large entourages of them along this year. Talking to attendees it seems this is a recurring theme this summer, the [SMBE \(Society for Molecular Biology and Evolution\) meeting in Ferrara](#) also on this week fully booked and there were some attendees who came to Lyon because they missed booking a place for the meeting in Italy. But all-in-all it was good to see people back in-person again and have the important experience of interacting and socialising with your scientific friends and peers.

GigaScience Macarons

Large Lyon Mischief

Our birthday celebrations were equally larger than expected, the Lyon-inspired GigaPanda macarons at the booth helping spread the word of the event at the Ninkasi microbrewery the other side of the LyonTech-la Doua University campus. Despite it being a bit of the walk, a public bus ran directly from the conference centre and at its peak these were being filled with party goers who arrived in bus-sized waves. The new *GigaByte* baby-GigaPanda (Pandalorian) t-shirts we handed out were popular with (human and non-human) attendees, but the 60 shirts we brought quickly ran out and we had over 100 attendees at the end of the night, even outnumbering our [legendary Prague 5th-birthday celebrations](#). You can see photos via our [facebook photo album](#) of the conference.

At the GigaScience party.

I, for one, applaud our AI-overlords

The recurring theme of the conference this was the rise of Large Language Models, and this topic infiltrated many of the COSI tracks and there was even a [Special Session](#) organized jointly by the ISCB Publication Committee and the ISCB Science in Society Committee on "Large Language Models – Are these the next pocket calculators?" that specifically covered the subject. This session included talks on the impact of LLMs on education, teaching and learning from Patricia Palagi; the ethical implications for science as a community of practice from David Leslie; and interestingly for us Alex Bateman giving an Editor's perspective of the challenges for journals having to deal with LLM-generated "plausible nonsense". Alex's talk presented [the latest guidance and policies](#) on the use of LLMs the ISCB has drafted in association with Editors from the ISCB journals *Bioinformatics* and *Bioinformatics Advances*. This policy outlines the permitted uses of LLMs in the publication process and highlights what is not permitted (with the caveat that in such a rapidly moving area these policies is likely to evolve as the tools improve and peoples level of acceptance changes). At GigaScience Press we are following the [COPE guidelines](#) on Authorship and AI tools, and are drafting similar detailed guidance for our authors and reviewers on this important and challenging area. Watch this space for further news and updates on this.

Looking at the other tracks, the [FUNCTION COSI](#) had talks on using **LLMs to** improve annotation, and the Text mining COSI having some talks on its use there as well. The panel with Larry Hunter, J. Harry Caufield and Chris Dallago (nvidea) on "Applications of ChatGPT and large language models in biology and medicine" being particularly popular and discussing wide variety of topics including trustworthiness and the need for better testing, assessment and training data.. It was not all negative, as Aurélie Névéol was in her Text Mining track talk on reproducibility noted that its increasingly important to show ChatGPT's limitations and this potentially increases the importance of sharing negative results and bringing them back in the game.

Back to BOSC

LLM's were even in our favourite [BOSC \(Bioinformatics Open Source Conference\) COSI](#), on top of AI-generated art prefacing the sessions, J. Harry Caufield presented an overview of transforming unstructured biomedical texts using this approach. Trying to fix "hallucinations" by grounding them in ontologies with the [OntoGPT tool](#). The two keynotes this year both covered very important subjects in Open Science. [Sara El-Gebali](#) opened the meeting presenting on the future of scientific progress through open collaboration. Busting the myths of scientific meritocracy, and telling us what we personally can do to address these issues. Including acknowledging bias, learn from mistakes, embrace diversity, recognise your role and privilege, and finally leveraging all of these to make better decision and change

Joseph Yracheta covered the dissonance between scientific altruism and capitalist extraction, and how indigenous communities need to look at federated data structures to better track the source and reuse of indigenous knowledge for benefit sharing and defining community

expectations. His [Native BioData Consortium](#) being an example of this approach as the only research institute and data repository hosted on indigenous land (domestic dependent nation) and led and run by indigenous scientists and tribal members in the US.

As always BOSC included talks from many of our long time collaborators, friends and authors. *GigaByte* Board Member Monica Munoz-Torres presented on the standards to connect biomedical data to AI in the [NIH Bridge2AI Program](#). There was also a talk from Tazro Ohta on [his recently published work](#) in *GigaScience* on a workflow reproducibility scale for automatic validation of biological interpretation results.

It was great to see many of our previous authors presenting. Amber Scholz in [the Human Frontier Science Program \(HFSP\) Symposium](#) presenting data from her "[myth busting](#)" paper in *GigaScience* on the current negotiations and potential effects of UN Convention on Biological Diversity on biological data management. It was standing room only for Matthieu Foll from IARC's Rare Cancer Genomics initiative keynote in VarI COSI on Multi-omics characterization of rare heterogeneous tumours. As an Open Science advocate using best practice in sharing workflows and interactive notebooks, the data underlying the mesothelioma research he presented here has all been described and shared and gathered together for reuse via a recent [GigaScience Data Note](#). We [previously published a Q&A](#) with Matthieu on his experience [sharing](#), accessing and reviewing controlled access data of rare cancers.

As in-person conferences are well and truly back we look forward to ISMB and BOSC 2024 in Montreal next year, and will make sure we book in sufficient enough advance to get a place as all the predictions are is it will be a large one.

The post [Going Large \(Language Models\) at ISMB2023](#) appeared first on [GigaBlog](#).